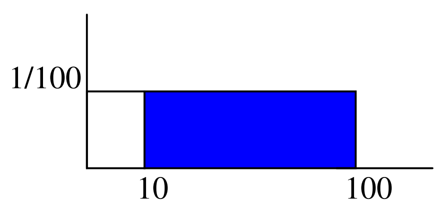
By Chieh Wu
Feb/13/2004

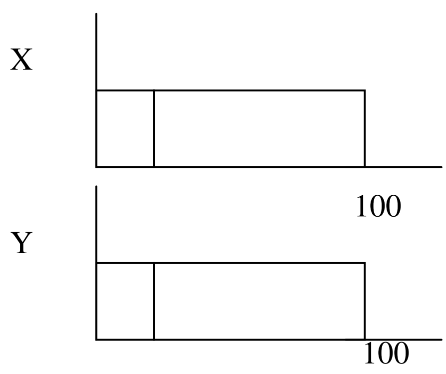
Problem:
Given 100 numbers that add up to a 100. The initial pdf of each RV is an uniform distribution range from 0<x<100. How are the numbers of N's you have chosen related to the certainty of obtaining a specific sum?

Let us start by viewing this question in a simpler perspective. Let's say if we only have one number, what is the probability that it would be over 10?

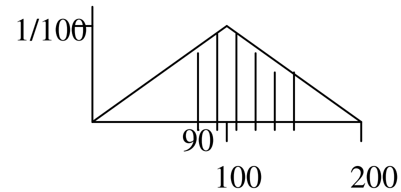


The result would be 90%.

What if we are using two? The point to notice here is that the combined pdf of two independent RV is the convolution of the functions.



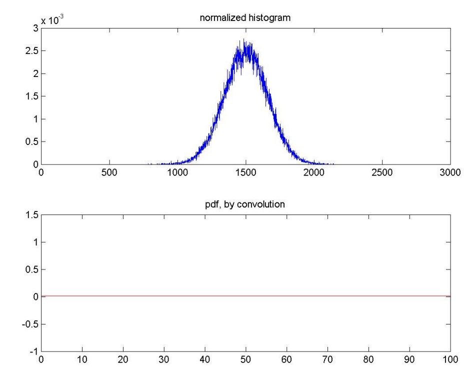
X + Y = *f*(X) * *f*(Y)



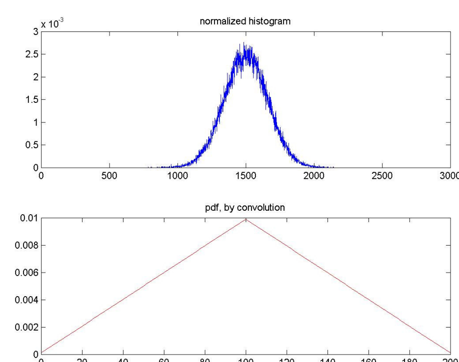
This is the pdf of two variables. We can find the probability of have a sum greater than 90 by finding the area above 90.

According to the central limit theory, given independent trials of N. As N approach infinity, the pdf of the combined RVs approaches a perfect normal distribution with *N(u,Var). u is the mean and Var is the variance.*

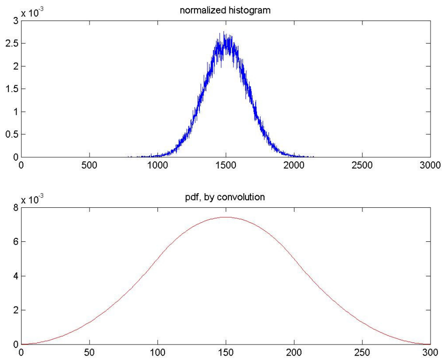
N = 1



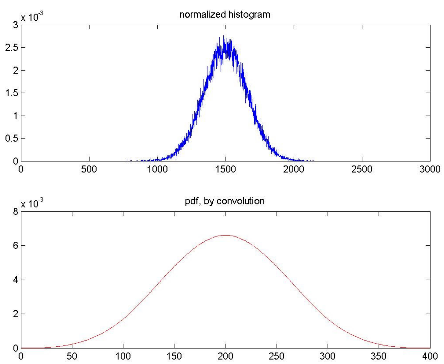
N = 2



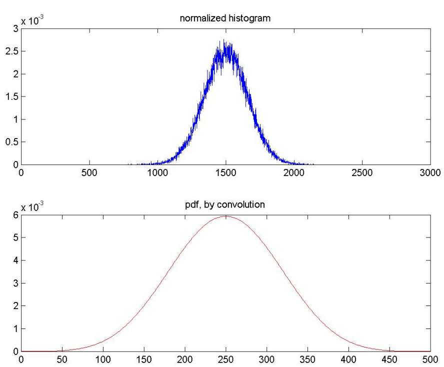
N = 3



N = 4



N > 4



As you can see, as N increases given that individual trials being independent, the graph also become more and more Gaussian. The mean of the Gaussian is the sum of individual mean and the variance is also the sum.

$$Y = a + b$$

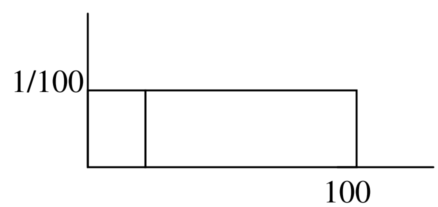
$$E[Y]= E[a] + E[b]$$

$$Var[Y] = Var[X] + Var[Y]$$

Unfortunately, this is not the case for this question. The fundamental premise of the previous assumption is the independence of individual trials. Although the pdf of individual trials might be identical as a initial condition, the pdf is actually a variable pdf that changes constantly depending on the previous results.

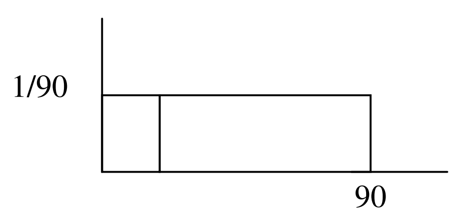
For example, let’s say we use only two RVs. The pdf of the first functions is the initial condition pdf.

N1

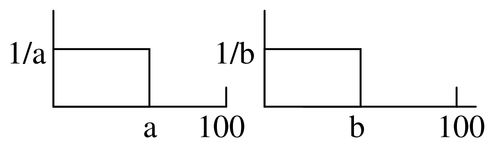


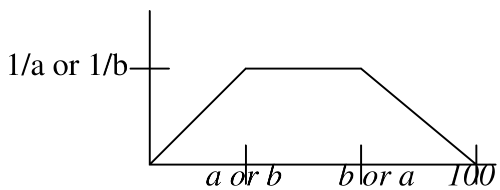
But let’s say the first number turn out to be 10. This immediately effects the pdf of the second variable because now it can only range from 0 to 90.

N2



The pdf of individual X_n is actually now a conditional probability that depends on all previous outcomes. This is a non-causal system. If we were to solve for the pdf of two variables symbolically we will notice that it does not provide more information.





From this point of view, it is theoretically impossible to determine the pdf without first perform the experiment.

However, there are some observational and experimental data we can draw from the total convolution of the sample.

1. Regardless of the sample number, the result pdf will always be a **symmetrical** function range from **0<X<100** with a mean at 50.

2. The increase of the sample decreases the variance.
   $Var = \Sigma Var_n$
   $Var = mean^2/12$

3. The area of the pdf is always 1.

The variance Some example of the convoluted
pdf